International Journal of Mathematics and Statistics Studies
Vol.11, No.2, pp.22-34, 2023
Print ISSN: 2053-2229 (Print),
Online ISSN: 2053-2210 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Performance Evaluation of Canonical Correlation Analysis and Generalized Canonical Correlation Analysis with Some Continuous Distributed Data

Okoli , C.N. and Eze-Golden, C.T
Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria

doi: https://doi.org/10.37745/ijmss.13/vol11n22234 Published August 20 2023

Citation: Okoli , C.N. and Eze-Golden, C.T (2023) Performance Evaluation of Canonical Correlation Analysis and Generalized Canonical Correlation Analysis with Some Continuous Distributed Data, *International Journal of Mathematics and Statistics Studies* Vol.11, No.2, pp.22-34

ABSTRACT: *This study was embarked to examine the performance evaluation of canonical correlation and generalized canonical correlation analysis with some continuous distributed data (Gamma, Gaussian, Exponential and Beta). The objectives of the study were to: ascertain if the anthropometric indicators of patients were correlated; ascertain if there is any relationship between vital signs and anthropometric dimensions of patients; obtain the relative efficiency of CCA and GCCA techniques for four continuous distributed simulated data; and determine the model performance adequacy of CCA and GCCA techniques. Real life medical data set was used, consisting of three response variables (Respiration rate, heart rate, temperature) named the vital signs and three predictor variables (Hip circumference, weight, height) named anthropometric dimensions. The study employed the real life data set to simulate data of sample sizes 15, 30, 45, 60, 100, 120, 140, 160, 400, 600, 800 and 1000 for the four continuous distributions. A computer programming language codes were written via R-Studio package to solve the numerous numerical problems in this study. The result of the study revealed that anthropometric dimensions, being the independent variables were not correlated, which implied that there was no symptom of multicollinearity using the Eigen values/condition index technique. In addition, there was significant relationship between vital signs and anthropometric dimensions of patients using Wilks' Lambda, Hotelling-Lawley Trace, Pillai's Bartlett Trace and Roy's Largest Root multivariate statistics. The adequacy of the CCA and GCCA was evaluated using Wilcoxon rank sum test; and the result revealed that GCCA was more efficient than that of CCA for the Gamma and Beta distributed data, while for Gaussian and Exponential distributed data, the relative efficiency of the CCA and GCCA was the same.*

KEYWORDS: Canonical correlation analysis, Generalized canonical correlation analysis, Gamma, Gaussian, Beta, Exponential, Simulated data.

INTRODUCTION

Canonical Correlation Analysis (CCA), proposed by Hotelling in 1936, is often employed for the study concerning interrelations between two sets of variables (Górecki et al, 2020). It consists of determining a linear transformation of the original variables from both sets into two new sets of variables not correlated within the sets but most highly correlated between them. Pairs of the corresponding new variables are called canonical variates, and the coefficients of correlation within the pairs are called canonical correlations.

Carroll in 1968 proposed a technique known as generalized canonical correlation analysis. In generalized canonical correlation analysis, $k > 2$ sets of variables are being analyzed simultaneously (Onyeagu et al., 2014). The central problem of GCCA is to construct a series of components aiming to maximize the association among the multiple variable sets. Although several generalizations of canonical correlation analysis have been proposed, some of which are discussed and compared in Kettenring and Gower in the years 1971 and 1989 respectively (Panzera et al., 2020). Carroll's approach has some attractive properties that make the method well suited to the analysis of multiple-set data (Van de Velden, 2011), which are as follows:
1. Computationally, the method is straightforward and its solution is based on an eigen-analysis.
2. The method is closely related to several well-known multivariate techniques such as principal component analysis, partial least squares and multivariate linear regression.
3. When the number of data sets $k = 2$, Carroll's GCCA reduces to the usual canonical correlation analysis.

This study therefore was aimed to: ascertain if the anthropometric indicators of patients were correlated; ascertain if there was any relationship between vital signs and anthropometric dimensions of patients; obtain the relative efficiency of CCA and GCCA techniques for four continuous distributed simulated data; and determine the model performance adequacy of CCA and GCCA techniques.

REVIEW OF RELATED LITERATURE

Makino (2022) worked on rotation in correspondence analysis from the canonical correlation perspective. According to the researcher, correspondence analysis (CA) is a statistical method for depicting the relationship between two categorical variables, and usually placed an emphasis on graphical representations. The study discussed a CA formulation based on canonical correlation analysis (CCA). In CCA-based formulation, the correlations within and between row/column categories in a reduced dimensional space could be expressed by canonical variables. However, in existing CCA-based formulations, only orthogonal rotation was permitted. They study proposed an alternative CCA-based formulation that permitted oblique rotation. In the proposed formulation, the CA loss function was defined as maximizing the generalized coefficient of determination, which was a measure of proximity between two variables. Simulation studies and real data examples were presented in order to demonstrate the benefits of the proposed formulation.

Nayir and Saridas (2022) worked on the relationship between culturally responsive teacher roles and innovative work behavior: Canonical correlation analysis. The aim of the study was to identify the relationship between culturally responsive teacher roles and innovative work behavior according to teachers' views. The first phase of the analysis revealed that in the first canonical function, which was calculated to maximize the relationship between culturally responsive teacher roles and innovative work behavior data sets, culturally responsive teacher roles and innovative work behavior data sets shared a variance of approximately 77%. Furthermore, as a result of the canonical correlation analysis, they determined that there was a positive relationship between the variables of the culturally regulating teacher (CRT) and the culturally mediating teacher (CMT) in the culturally responsive teacher roles data set and the GII and FSI variables in the innovative work behavior data set.

McKeague and Zhang (2021) researched on the significance testing for canonical correlation analysis in high dimensions. They considered the problem of testing the presence of a relationship which was linear between large sets of random variables based on a post-selection inference technique to canonical correlation analysis. The adjustment for the selection of subsets of variables having linear combinations with maximal sample correlation was a challenge. In that regard, they constructed a stabilized on-step estimator of the Euclidean norm of the canonical correlations maximized over subsets of variables of pre-specified cardinality. The estimator was shown to be consistent for its target parameter and asymptotically normal, provided the dimensions of the variables did not grow too quickly with the sample size. The authors developed a greedy search algorithm to accurately compute the estimator, leading to a computationally tractable omnibus test for the global null hypothesis that there were no linear associations between any subsets of variables having the pre-specified cardinality. In addition, they developed a confidence interval that took the variable selection into consideration.

MATERIALS AND METHODS

Canonical Variates and Canonical Correlations

The canonical correlations measure the strength of association between the two sets of variables. The first group of p variables is represented by the $(p \times 1)$ random vector $\mathbf{X}^{(1)}$, while the second group of q variables is represented by the $(q \times 1)$ random vector $\mathbf{X}^{(2)}$. It will be assumed, in the theoretical development, that $\mathbf{X}^{(1)}$ represents the smaller set, so that $p \leq q$.
For the random vectors $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, let

$$\left.\begin{array}{ll} E(\mathbf{X}^{(1)})=\boldsymbol{\mu}^{(1)}; & Cov(\mathbf{X}^{(1)})=\Sigma_{11} \\ E(\mathbf{X}^{(2)})=\boldsymbol{\mu}^{(2)}; & Cov(\mathbf{X}^{(2)})=\Sigma_{22} \\ Cov(\mathbf{X}^{(1)},\mathbf{X}^{(2)})=\Sigma_{11}=\Sigma_{22}' & \end{array}\right\} \quad (1)$$

It will be convenient to consider $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ jointly, so, the random vector

$$\underset{((p+q)\times 1)}{\mathbf{X}}=\begin{bmatrix}\mathbf{X}^{(1)}\\ \hdashline \mathbf{X}^{(2)}\end{bmatrix}=\begin{bmatrix}X_1^{(1)}\\ X_2^{(1)}\\ \vdots\\ X_p^{(1)}\\ \hdashline X_1^{(2)}\\ X_2^{(2)}\\ \vdots\\ X_q^{(2)}\end{bmatrix} \quad (2)$$

has mean vector

$$\underset{((p+q)\times 1)}{\boldsymbol{\mu}}=E(\mathbf{X})=\begin{bmatrix}E(\mathbf{X}^{(1)})\\ \hdashline E(\mathbf{X}^{(2)})\end{bmatrix}=\begin{bmatrix}\boldsymbol{\mu}^{(1)}\\ \hdashline \boldsymbol{\mu}^{(2)}\end{bmatrix} \quad (3)$$

and covariance matrix

$$\underset{(p+q)\times(p+q)}{\boldsymbol{\Sigma}}=\mathrm{E}(\mathbf{X}-\boldsymbol{\mu})\mathrm{E}(\mathbf{X}-\boldsymbol{\mu})'$$

$$=\left[\begin{array}{c:c} E(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})' \\ \hdashline E(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\mathbf{X}^{(1)}-\boldsymbol{\mu}^{(1)})' & E(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})(\mathbf{X}^{(2)}-\boldsymbol{\mu}^{(2)})' \end{array}\right]$$

$$\left[\begin{array}{c:c} \underset{(p\times p)}{\boldsymbol{\Sigma}_{11}} & \underset{(p\times q)}{\boldsymbol{\Sigma}_{12}} \\ \hdashline \underset{(q\times p)}{\boldsymbol{\Sigma}_{21}} & \underset{(q\times q)}{\boldsymbol{\Sigma}_{22}} \end{array}\right] \quad (4)$$

Generalized Canonical Correlation Analysis

Roy has proposed formal generalization of the notion of canonical correlation to three or more sets of variates (Van de Veldon, 2011). Considering n column centered observations X_i of order $m\times p_i$, then the generalized canonical correlation is given as

$$\left.\begin{array}{l} \min_{Y,A_i}\ \theta=trace\sum_{i=1}^{n}\left(Y-X_iA_i\right)'\left(Y-X_iA_i\right) \\ \qquad S.t.\quad Y'Y=I_k \end{array}\right\} \quad (5)$$

Where Y is the group configuration matrix

It has be shown that the matrices A_i can be computed as according to Carol in 1968 as (Onyeagu et al, 2014)

$$A_i = (X_i' X_i)^{-1} X_i' Y \tag{6}$$

The group configuration matrix y can be computed using eigen equation as

$$\sum_{i=1}^{n} (X_i (X_i' X_i)^{-1} X_i') Y = Y\Lambda \tag{7}$$

where Λ is a diagonal matrix with diagonal elements y_j, being the k large eigenvalue of $\sum_{i=1}^{n} (X_i (X_i' X_i)^{-1} X_i'),\ X_{i's}'$ assumed full column rank and the column of Y are corresponding eigenvectors.

Tests of Significance in CCA

Wilk's Lambda Test Statistic

The null hypothesis is the same with the statement that all canonical correlations ($r_1, r_2, \cdots, r_s$) are non-significant (Alpar, 2013). The significance of $r_1, r_2, \cdots, r_s$ can be tested by

$$\Lambda_1 = \frac{|S|}{|S_{yy}||S_{zz}|} = \frac{|R|}{|R_{yy}||R_{zz}|}, \tag{8}$$

Which is distributed as $\Lambda_{p,q,n-1-q}$. H$_0$ is rejected if $\Lambda_1 \le \Lambda_\alpha$, where Λ_α are the critical values available in Statistical Table by employing $v_H = q$ and $v_E = n-1-q$. The statistic Λ_1 in Equation (8) is also distributed as $\Lambda_{q,p,n-1-q}$. The statistic Λ_1 can be expressed in terms of the squared canonical correlations:

$$\Lambda_1 = \prod_{i=1}^{s} (1 - r_i^2) \tag{9}$$

However, if the parameters exceed the range of critical values for Wilks' $\wedge$ in the statistical Table, the chi-square approximation can be employed as;

$$\chi^2 = -\left[n - \frac{1}{2}(p+q+3)\right] \ln \Lambda_1, \tag{10}$$

which is approximately distributed as chi-square distribution with *pq* degrees of freedom. In this case, H$_0$ is rejected if chi-square statistic is greater than or equal to chi-square tabulated.

Pillai's Test Statistic

Pillai's test statistic for the significance of canonical correlations is

$$V^{(s)} = \sum_{i=1}^{s} r_i^2 \tag{11}$$

The upper percentage points of $V^{(s)}$ are found in the Statistical Table, which is indexed by

$s = \min(p, q)$, $m = \frac{1}{2}(|q - p| - 1)$, $N = \frac{1}{2}(n - q - p - 2)$.

Lawley–Hotelling Statistic

The Lawley–Hotelling statistic for canonical correlations is

$$U^{(s)} = \sum_{i=1}^{s} \frac{r_i^2}{1 - r_i^2} \tag{12}$$

Upper percentage points for $\frac{v_E U^{(s)}}{v_H}$ are given in Statistical Table, which is entered with *p*, $v_H = q$, and $v_E = n - q - 1$.

Roy's Largest Root Statistic

Roy's largest root statistic is given by

$$\theta = r_1^2 \tag{13}$$

The upper percentage points are gotten in Statistical Table, with s, m, and N defined in Pillai's test statistic.

RESULTS OF DATA ANALYSIS

Test to ascertain if the Anthropometric Dimensions of Patients were correlated

Table 4.1: Eigenvalues and Condition Index

Variables	Eigenvalue	Condition Index
Hip Circumference	0.009	21.489
Weight	0.006	26.237
Height	0.001	87.805

Since CI = 3 < 10, there is no significant multicollinearity. Thus, this implied no multicollinearity symptoms.

Test to ascertain if there was any Relationship between Vital Signs and Anthropometric Dimensions of Patients

Table 4.2: Multivariate Test Statistics

Statistic	Stat	Approximate F	df1	df2	p-value
Wilks' Lambda	0.16174	58.45691	9	472	0.000**
Hotelling-Lawley Trace	**5.00795**	**79.51741**	**9**	**578**	**0.000****
Pillai's Bartlett Trace	0.86656	26.53719	9	588	0.000**
Roy's Largest Root	0.83257	324.88750	3	196	0.000**

Footnote: **=Sig. at 5%

The results in Table 4. 2 showed that all the multivariate test statistics were significant at 5% level of significance. This concluded that there was significant relationship between vital signs and anthropometric dimensions of patients.

To Analyze the Simulated Data of Different Sample Sizes for CCA and GCCA

Data were simulated on R-Studio command window, calling for the CCA and GCCA function for Gamma distribution, Gaussian distribution, Exponential and the Beta distribution for samples of sizes 15, 30, 45, 60, 100, 120, 140, 160, 400, 600, 800 and 1000, and the results obtained are summarized in Table 4.3

Table 4. 3: Summary Results from the Four Distributions for Different Sample Sizes

Distribution		Correlation	Eigen-value	X-Mean Vector		Y-Mean Vector	
	Sample	CCA	GCCA	CCA	GCCA	CCA	GCCA
		0.7331	0.8665	110.2051	110.2051	17.8201	17.8201
Gamma	15	0.3499	0.6750	70.9946	70.9946	38.9106	38.9106
		0.0576	0.5288	167.3032	167.3032	62.9430	62.9430
		SD = 0.3388	**SD = 0.1694**				
		0.5643	0.7822	108.5295	108.5295	17.5992	17.5992
	30	0.2555	0.6277	68.9724	68.9724	38.4738	38.4738
		0.0288	0.5144	169.0976	169.0976	62.2536	62.2536
		SD = 0.2688	**SD = 0.1344**				
		0.2920	0.6460	106.3649	106.3649	18.1204	18.1204
	45	0.2248	0.6124	70.3873	70.3873	38.9415	38.9415
		0.1243	0.5622	166.0031	166.0031	60.4868	60.4868
		SD = 0.0844	**SD = 0.0422**				
		0.3789	0.6895	105.5420	105.5420	17.8711	17.8711
	60	0.1698	0.5849	69.8049	69.8049	38.8489	38.8489
		0.0642	0.5321	167.5262	167.5262	61.0660	61.0660
		SD = 0.1602	**SD = 0.0801**				
		0.3002	0.6501	105.1322	105.1322	18.0607	18.0607
	100	0.1603	0.5802	69.6309	69.6309	38.7851	38.7851
		0.0860	0.5430	167.7657	167.7657	61.3914	61.3914
		SD = 0.1088	**SD = 0.0544**				
		0.1909	0.5954	104.9475	104.9475	18.1063	18.1063
	120	0.1349	0.5674	69.1980	69.1980	38.8651	38.8651
		0.0684	0.5342	168.0896	168.0896	61.5755	61.5755
		SD = 0.0613	**SD = 0.0306**				

		0.2663	0.6331	103.9614	103.9614	17.9772	17.9772
	140	0.1322	0.5661	69.5637	69.5637	38.9455	38.9455
		0.0405	0.5203	168.0234	168.0234	61.0524	61.0524
		SD = 0.1136	**SD = 0.0567**				
		0.1741	0.5870	104.0724	104.0724	18.3032	18.3032
	160	0.1377	0.5688	69.8922	69.8922	38.7852	38.7852
		0.0728	0.5364	167.6923	167.6923	61.0048	61.0048
		SD = 0.0513	**SD = 0.0256**				
		0.1745	0.5873	104.1268	104.1268	17.8753	17.8753
	400	0.0698	0.5349	69.9637	69.9637	38.8092	38.8092
		0.0318	0.5159	168.0265	168.0265	61.1837	61.1837
		SD = 0.0739	**SD = 0.0370**				
		0.0970	0.5485	104.3280	104.3280	18.0180	18.0180
	600	0.0494	0.5247	69.8431	69.8431	38.7632	38.7632
		0.0047	0.5023	167.8896	167.8896	61.1633	61.1633
		SD = 0.0462	**SD = 0.0231**				
		0.0964	0.5482	104.4540	104.4540	17.9205	17.9205
	800	0.0534	0.5267	69.5324	69.5324	38.7973	38.7973
		0.0157	0.5078	168.0250	168.0250	61.1682	61.1682
		SD = 0.0404	**SD = 0.0202**				
		0.3002	0.6501	105.1322	105.1322	18.0607	18.0607
	1000	0.1603	0.5802	69.6309	69.6309	38.7851	38.7851
		0.0860	0.5430	167.7657	167.7657	61.3914	61.3914
		SD = 0.1088	**SD = 0.0544**				
Gaussian		0.5852	0.7926	99.1845	99.1845	18.8663	18.8663
	15	0.3826	0.6913	70.8426	70.8426	38.9333	38.9333
		0.3025	0.6513	167.6035	167.6035	65.3529	65.3529
		SD = 0.1457	**SD = 0.0728**				
		0.4997	0.7499	102.9485	102.9485	18.1755	18.1755
	30	0.3696	0.6848	70.5553	70.5553	38.8055	38.8055
		0.0026	0.5013	170.1050	170.1050	57.5391	57.5391
		SD = 0.2578	**SD = 0.1289**				
		0.4113	0.7057	103.0802	103.0802	17.2762	17.2762
	45	0.3403	0.6701	72.1554	72.1554	38.7726	38.7726
		0.0412	0.5206	167.7889	167.7889	61.6741	61.6741
		SD = 0.1964	**SD = 0.0982**				
		0.3160	0.6580	104.5254	104.5254	18.0498	18.0498
	60	0.1517	0.5758	71.1767	71.1767	38.7192	38.7192
		0.0784	0.5392	166.4579	166.4579	60.1706	60.1706
		SD = 0.1217	**SD = 0.0608**				
		0.3341	0.6671	106.5264	106.5264	17.9504	17.9504
	100	0.1614	0.5807	68.5013	68.5013	38.8288	38.8288
		0.0330	0.5165	167.7350	167.7350	61.5448	61.5448
		SD = 0.1511	**SD = 0.0756**				
		0.2174	0.6087	105.9736	105.9736	18.1295	18.1295
	120	0.1224	0.5612	68.7890	68.7890	38.8742	38.8742
		0.0312	0.5156	167.3116	167.3116	60.6445	60.6445
		SD = 0.0931	**SD = 0.0465**				
		0.2612	0.6306	105.4082	105.4082	18.1290	18.1290

	140	0.1721	0.5860	69.1872	69.1872	38.7523	38.7523
		0.0462	0.5231	167.6334	167.6334	61.9933	61.9933
		SD = 0.1080	**SD = 0.0540**				
		0.1384	0.5692	104.8525	104.8525	18.1566	18.1566
	160	0.1190	0.5595	68.9627	68.9627	38.8621	38.8621
		0.0374	0.5187	167.9299	167.9299	61.5277	61.5277
		SD = 0.0536	**SD = 0.0268**				
		0.1410	0.5705	103.9314	103.9314	18.0633	18.0633
	400	0.0675	0.5338	69.9915	69.9915	38.8047	38.8047
		0.0331	0.5166	168.2507	168.2507	61.4925	61.4925
		SD = 0.0551	**SD = 0.0275**				
		0.0689	0.5344	104.2548	104.2548	18.0775	18.0775
	600	0.0498	0.5249	70.0278	70.0278	38.7873	38.7873
		0.0238	0.5119	168.1172	168.1172	61.6193	61.6193
		SD = 0.0226	**SD = 0.0113**				
		0.1053	0.5526	104.4183	104.4183	18.0523	18.0523
	800	0.0340	0.5170	69.8996	69.8996	38.7953	38.7953
		0.0073	0.5036	168.1137	168.1137	61.6126	61.6126
		SD = 0.0507	**SD = 0.0253**				
		0.0785	0.5392	104.4183	104.4183	18.0523	18.0523
	1000	0.0416	0.5308	69.8996	69.8996	38.7953	38.7953
		0.0085	0.5042	168.1137	168.1137	61.6126	61.6126
		SD = 0.0350	**SD = 0.0183**				
Exponential		0.5639	0.7820	46.8736	46.8736	21.7214	21.7214
	15	0.2391	0.6196	55.8412	55.8412	34.4638	34.4638
		0.1752	0.5876	115.9493	115.9493	71.9740	71.9740
		SD = 0.2084	**SD = 0.1042**				
		0.5834	0.7917	115.2464	115.2464	14.9528	14.9528
	30	0.4403	0.7201	66.0048	66.0048	38.0459	38.0459
		0.0943	0.5472	173.1484	173.1484	66.2511	66.2511
		SD = 0.2515	**SD = 0.1257**				
		0.6093	0.8047	100.8783	100.8783	20.7309	20.7309
	45	0.0889	0.5445	75.9062	75.9062	46.6858	46.6858
		0.0002	0.5001	130.7947	130.7947	55.3866	55.3866
		SD = 0.3291	**SD = 0.1645**				
		0.2677	0.6339	107.0425	107.0425	20.2386	20.2386
	60	0.0852	0.5426	64.8264	64.8264	32.7580	32.7580
		0.0423	0.5211	173.0728	173.0728	53.8399	53.8399
		SD = 0.1197	**SD = 0.0599**				
		0.1762	0.5881	106.7329	106.7329	18.7833	18.7833
	100	0.0707	0.5353	71.5469	71.5469	37.4559	37.4559
		0.0029	0.5015	154.8657	154.8657	63.8125	63.8125
		SD = 0.0873	**SD = 0.0436**				
		0.1792	0.5896	102.0583	102.0583	19.3451	19.3451
	120	0.1139	0.5569	75.0108	75.0108	39.8900	39.8900
		0.0608	0.5304	144.6335	144.6335	61.3095	61.3095
		SD = 0.0593	**SD = 0.0297**				
		0.3888	0.6944	98.1591	98.1591	17.5587	17.5587
	140	0.1346	0.5673	76.2943	76.2943	40.2088	40.2088

		0.0328	0.5164	160.7795	160.7795	61.5865	61.5865
		SD = 0.1834	**SD = 0.0917**				
		0.2026	0.6013	102.6147	102.6147	18.7412	18.7412
	160	0.1792	0.5896	67.2862	67.2862	36.5142	36.5142
		0.0036	0.5018	174.7996	174.7996	69.6896	69.6896
		SD = 0.1088	**SD = 0.0546**				
		0.1344	0.5672	104.6953	104.6953	18.3324	18.3324
	400	0.0796	0.5398	68.6779	68.6779	39.3008	39.3008
		0.0419	0.5209	166.0078	166.0078	58.1190	58.1190
		SD = 0.0465	**SD = 0.0233**				
		0.1122	0.5561	104.7341	104.7341	17.7330	17.7330
	600	0.0708	0.5354	68.4380	68.4380	37.6497	37.6497
		0.0059	0.5030	168.3519	168.3519	62.7830	62.7830
		SD = 0.0536	**SD = 0.0268**				
		0.0743	0.5371	103.7684	103.7684	17.5385	17.5385
	800	0.0514	0.5257	69.8646	69.8646	39.1916	39.1916
		0.0192	0.5096	164.6596	164.6596	63.8313	63.8313
		SD = 0.0277	**SD = 0.0138**				
		0.0563	0.5281	106.3430	106.3430	18.1171	18.1171
	1000	0.0302	0.5151	68.6035	68.6035	40.2404	40.2404
		0.0215	0.5107	161.4812	161.4812	63.8001	63.8001
		SD = 0.0181	**SD = 0.0090**				
Beta		0.6297	0.8148	0.5035	0.5035	0.5109	0.5109
	15	0.5965	0.7982	0.5028	0.5028	0.5059	0.5059
		0.2519	0.6260	0.5015	0.5015	0.5025	0.5025
		SD = 0.2092	**SD = 0.1045**				
		0.3903	0.6952	0.5025	0.5025	0.5131	0.5131
	30	0.2115	0.6058	0.5031	0.5031	0.5071	0.5071
		0.0320	0.5160	0.5011	0.5011	0.5049	0.5049
		SD = 0.1792	**SD = 0.0896**				
		0.5391	0.7695	0.5025	0.5025	0.5191	0.5191
	45	0.3763	0.6881	0.5022	0.5022	0.5065	0.5065
		0.1069	0.5534	0.5014	0.5014	0.5049	0.5049
		SD = 0.2183	**SD = 0.1091**				
		0.3596	0.6798	0.5022	0.5022	0.5152	0.5152
	60	0.1733	0.5866	0.5026	0.5026	0.5067	0.5067
		0.0738	0.5369	0.5017	0.5017	0.5040	0.5040
		SD = 0.1451	**SD = 0.0725**				
		0.2504	0.6252	0.5016	0.5016	0.5139	0.5139
	100	0.1099	0.5550	0.5041	0.5041	0.5065	0.5065
		0.0141	0.5071	0.5016	0.5016	0.5038	0.5038
		SD = 0.1189	**SD = 0.0594**				
		0.2736	0.6368	0.5017	0.5017	0.5146	0.5146
	120	0.1486	0.5743	0.5042	0.5042	0.5064	0.5064
		0.0008	0.5004	0.5015	0.5015	0.5038	0.5038
		SD = 0.1366	**SD = 0.0682**				
		0.2253	0.6127	0.5018	0.5018	0.5135	0.5135
	140	0.1376	0.5688	0.5043	0.5043	0.5065	0.5065
		0.0081	0.5040	0.5015	0.5015	0.5041	0.5041

	SD = 0.1093	**SD = 0.0547**				
	0.1376	0.5688	0.5024	0.5024	0.5137	0.5137
160	0.1013	0.5507	0.5038	0.5038	0.5064	0.5064
	0.0368	0.5184	0.5015	0.5015	0.5040	0.5040
	SD = 0.0511	**SD = 0.0255**				
	0.1031	0.5516	0.5025	0.5025	0.5143	0.5143
400	0.0755	0.5377	0.5035	0.5035	0.5066	0.5066
	0.0324	0.5162	0.5015	0.5015	0.5038	0.5038
	SD = 0.0356	**SD = 0.0178**				
	0.1062	0.5531	0.5024	0.5024	0.5139	0.5139
600	0.0750	0.5375	0.5035	0.5035	0.5066	0.5066
	0.0022	0.5011	0.5015	0.5015	0.5039	0.5039
	SD = 0.0534	**SD = 0.0267**				
	0.0773	0.5386	0.5024	0.5024	0.5142	0.5142
800	0.0512	0.5256	0.5036	0.5036	0.5065	0.5065
	0.0011	0.5005	0.5015	0.5015	0.5042	0.5042
	SD = 0.0387	**SD = 0.0194**				
	0.0721	0.5360	0.5023	0.5023	0.5139	0.5139
1000	0.0693	0.5347	0.5037	0.5037	0.5066	0.5066
	0.0008	0.5004	0.5015	0.5015	0.5040	0.5040
	SD = 0.0404	**SD = 0.0202**				

Table 4.3 shows the standard deviation of the correlations and eigenvalues for CCA and GCCA respectively. It can be observed that the standard deviation of the GCCA is lower than that of CCA, but there is need to examine if the difference is significant. It is also observed that the X and Y-variates of the CCA and GCCA do not differ.

Model Performance Adequacy of CCA and GCCA Techniques

Table 4.4: Summary of Decision for Testing SD Values for CCA and GCCA

		SD Values		**Ranks**		**Z**	**Decision**
Distribution	**Sample**	**CCA**	**GCCA**	**CCA**	**GCCA**		
	15	0.3388	0.1694	24	22		
	30	0.2688	0.1344	23	20		
	45	0.0844	0.0422	16	7		
	60	0.1602	0.0801	21	15		
	100	0.1088	0.0544	17.5	10.5		
Gamma	120	0.0613	0.0306	13	4	2.17	Reject H_0
	140	0.1136	0.0567	19	12		
	160	0.0513	0.0256	9	3		
	400	0.0739	0.0370	14	5		
	600	0.0462	0.0231	8	2		
	800	0.0404	0.0202	6	1		
	1000	0.1088	0.0544	17.5	10.5		
	15	0.1457	0.0728	21	14		
	30	0.2578	0.1289	24	20		
	45	0.1964	0.0982	23	17		

	60	0.1217	0.0608	19	13		
	100	0.1511	0.0756	22	15		
	120	0.0931	0.0465	16	8	1.93	Do not Reject H_0
Gaussian	140	0.1080	0.0540	18	11		
	160	0.0536	0.0268	10	5		
	400	0.0551	0.0275	12	6		
	600	0.0226	0.0113	3	1		
	800	0.0507	0.0253	9	4		
	1000	0.0350	0.0183	7	2		
	15	0.2084	0.1042	22	16		
	30	0.2515	0.1257	23	19		
	45	0.3291	0.1645	24	20		
	60	0.1197	0.0599	18	13		
	100	0.0873	0.0436	14	8		
	120	0.0593	0.0297	12	7	1.65	Do not Reject H_0
Exponential	140	0.1834	0.0917	21	15		
	160	0.1088	0.0546	17	11		
	400	0.0465	0.0233	9	4		
	600	0.0536	0.0268	10	5		
	800	0.0277	0.0138	6	2		
	1000	0.0181	0.0090	3	1		
	15	0.2092	0.1045	23	16		
	30	0.1792	0.0896	22	15		
	45	0.2183	0.1091	24	17		
	60	0.1451	0.0725	21	14		
	100	0.1189	0.0594	19	12		
	120	0.1366	0.0682	20	13	2.11	Reject H_0
Beta	140	0.1093	0.0547	18	11		
	160	0.0511	0.0255	9	4		
	400	0.0356	0.0178	6	1		
	600	0.0534	0.0267	10	5		
	800	0.0387	0.0194	7	2		
	1000	0.0404	0.0202	8	3		

Table 4. 4 shows the Wilcoxon rank sum test significance difference result for the four continuous distributions employed in this study. The result reveals that there is no significant difference in the standard deviation of the correlations and eigenvalues for the methods for Gaussian and Exponential distributions. This implies that the relative efficiency of the CCA and GCCA is the same for the Gaussian and Exponential distributed data. On the other hand, the result reveals that there is significant difference in the standard deviation of the correlations and eigenvalues for the methods for Gamma and Beta distributions. This implies that GCCA is more efficient than that of CCA for the Gamma and Beta distributed data.

CONCLUSION

This study used canonical correlation and generalized canonical correlation analysis using four continuous distributions (Gamma, Gaussian, Exponential and Beta) in order to assess their performances. The study concluded that the independent variables, which is the anthropometric dimensions were not correlated, that is there is no symptom of multicollinearity using the Eigen values/condition index technique. In addition, there is significant relationship between vital signs and anthropometric dimensions of patients using the four multivariate methods.

The adequacy of the CCA and GCCA was evaluated with Wilcoxon rank sum test; and the study concluded that GCCA is more efficient than that of CCA for the Gamma and Beta distributed data, while for Gaussian and Exponential distributed data, the relative efficiency of the CCA and GCCA is the same.

REFERENCES

Alpar, R. (2013). Applied multivariate statistical methods. Detay Publishing, Ankara, Turkey.

Góreck, T., Krzysko, M. & Wołynski, W. (2020). Generalized canonical correlation analysis for functional data. *Biometrical Letters*, *57*(2020), 1 – 12.

Makino, N. (2022). Rotation in correspondence analysis from the canonical correlation perspective. *Psychometrika,5(2022)*, 18–28.

McKeague, I. W. & Zhang, X. (2021). Significance testing for canonical correlation analysis in high dimensions. *Biometrika*, 2021.

Nayir, F. & Saridas, G. (2022). The relationship between culturally responsive teacher roles and innovative work behavior: canonical correlation analysis. *Journal of Educational Research and Practice*, *12(2022)*, 36 –50.

Onyeagu, S. I., Osuji, G. A. & Onyia, O. M. (2014). Comparison of canonical correlation analysis and the generalized canonical correlation analysis using the lognormal and Cauchy distributed data. *Mathematical Theory and Modeling*, *4*(5), 177 – 185.

Panzera, F., Bergamo, P. & Fäh, D. (2021). Canonical correlation analysis based on site response proxies to predict site specific amplification functions in Switzerland. Bulletin of the Seismological Society of America, *111(4),* 1905–1920.

Van de Velden. M. (2011). On generalized canonical correlation analysis. Proc. 58th World Statistical Congress. Dublin, 758–765.